
Topology Adaptive Graph Estimation in High Dimensions

Johannes Lederer
Cornell University

Christian L. Müller
Simons Center for Data Analysis, Simons Foundation

Abstract

We introduce Graphical TREX (GTREX), a novel method for graph estimation in high-dimensional Gaussian graphical models. By conducting neighborhood selection with TREX, GTREX avoids tuning parameters and is adaptive to the graph topology. We compare GTREX with standard methods on a new simulation set-up that is designed to assess accurately the strengths and shortcomings of different methods. These simulations show that a neighborhood selection scheme based on Lasso and an optimal (in practice unknown) tuning parameter outperforms other standard methods over a large spectrum of scenarios. Moreover, we show that GTREX can rival this scheme and, therefore, can provide competitive graph estimation without the need for tuning parameter calibration.

1 INTRODUCTION

Graphical models [Lauritzen, 1996] have become an important tool to find and describe patterns in high-dimensional data. In biology, for example, graphical models have been successfully applied to estimate interactions between genes from high-throughput expression profiles [Wille et al., 2004, Friedman, 2004], to predict contacts between protein residues from multiple sequence alignments [Jones et al., 2012], and to uncover interactions of microbes from gene sequencing data [Kurtz et al., 2014]. Graphical models represent the conditional dependence structure of the underlying random variables as a graph. Learning a graphical model from data requires a simultaneous estimation of the graph and of the probability distribution that factorizes according to this graph. In the Gaussian case, it is well known that the underlying graph

is determined by the non-zero entries of the precision matrix (the inverse of the population covariance matrix). Gaussian graphical models have become particularly popular after the advent of computationally efficient approaches, such as neighborhood selection [Meinshausen and Bühlmann, 2006] and sparse covariance estimation [Banerjee et al., 2008, Yuan and Lin, 2007], that can learn even high-dimensional graphical models. Neighborhood selection, on the one hand, reconstructs the graph by estimating the local neighborhood of each node via the Lasso [Tibshirani, 1996]. This approach is usually seen as a proxy to the covariance selection problem [Friedman et al., 2008]. On the other hand, [Banerjee et al., 2008] and [Yuan and Lin, 2007] showed that the graph and the precision matrix can be simultaneously estimated by solving a global optimization problem. State-of-the-art solvers are the Graphical Lasso [Friedman et al., 2008] and the Quadratic Approximation of Inverse Covariance (QUIC) method [Hsieh et al., 2011]. Both approaches can be extended beyond the framework of Gaussian graphical models. To mention two of the many examples, [Ravikumar et al., 2010] study neighborhood selection for Ising models, and [Liu et al., 2009] introduce a semi-parametric penalized likelihood estimator that allows for non-Gaussian distributions of the data.

Although the field has advanced tremendously in the past decade, there are still a number of challenges, both from a practical and a theoretical point of view. First, the conditions that are currently imposed [Meinshausen and Bühlmann, 2006, Ravikumar et al., 2010, Lam and Fan, 2009, Ravikumar et al., 2011] to show consistency in graph and/or graphical model estimation are difficult to meet or verify in practice. Moreover, the performance of any of the standard methods heavily depends on the simulation set-up or the data at hand [Liu and Ihler, 2011, Liu and Wang, 2012, Kurtz et al., 2014]. Furthermore, standard neighborhood selection and covariance estimation methods require a careful calibration of a tuning parameter, especially because the model complexity is known *a priori* only in very few examples [Jones et al., 2012]. In practice, the tuning parameters are calibrated via cross-validation, classical information criteria such as AIC

and BIC [Yuan and Lin, 2007], or stability criteria [Liu et al., 2010]. However, different calibration schemes can result in largely disparate estimates [Liu et al., 2010].

To approach some of the practical challenges, we introduce *Graphical TREX (GTREX)*, a novel method for graph estimation based on neighborhood selection with TREX [Lederer and Müller, 2014]. The main feature of GTREX is that it can make tuning parameters superfluous, which renders this method particularly useful in practice. We also introduce a novel simulation set-up that may serve as a benchmark to assess the strengths and shortcomings of different methods.

Our simulations show that, if the tuning parameter is optimally chosen, standard neighborhood selection with the “or-rule” outmatches other standard methods across a wide range of scenarios. Our simulations also show that GTREX can rival this method in many scenarios. Since optimal tuning parameters depend on unknown quantities and, therefore, are not accessible in practice, this demonstrates that GTREX is a promising alternative for graph estimation in high-dimensional graphical models.

The remainder of the paper is structured as follows. After specifying the framework and notation, we introduce GTREX in Section 2. We then describe the experimental scenarios in Section 3 and present the numerical results in Section 4. We finally conclude in Section 5.

FRAMEWORK AND NOTATION

We consider n samples from a p -dimensional Gaussian distribution $\mathcal{N}_p(0, \Sigma)$ with positive definite, symmetric covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$. The samples are summarized in the matrix $X \in \mathbb{R}^{n \times p}$ such that X_{ij} corresponds to the j th component of the i th sample. We call Σ^{-1} the precision matrix and note that the precision matrix is symmetric.

The Gaussian distribution $\mathcal{N}_p(0, \Sigma)$ can be associated with an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, p\}$ is the set of nodes and $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ the set of (undirected) edges that consists of all pairs $(i, j), (j, i) \in \mathcal{V} \times \mathcal{V}$ that fulfill $i \neq j$ and $(\Sigma^{-1})_{ij} \neq 0$. We denote by e_{ij} (and equivalently by e_{ji}) the edge that corresponds to the pair $(i, j), (j, i)$ and by $k := |\mathcal{E}|$ the number of edges in the graph \mathcal{G} .

We denote by $\text{supp}(\beta)$ the support of a vector β , by $a \vee b$ and $a \wedge b$ the maximum and minimum, respectively, of two constants $a, b \in \mathbb{R}$, and by $|\cdot|$ the cardinality of a set.

In this paper, we focus on estimating which entries of the precision matrix Σ^{-1} are non-zero from the data X .

This is equivalent to estimating the set of edges \mathcal{E} from X . We assess the quality of an estimate $\hat{\mathcal{E}}$ via the Hamming distance to the true set of edges \mathcal{E} given by $d_H(\hat{\mathcal{E}}, \mathcal{E}) := |\{e_{ij} : e_{ij} \in \hat{\mathcal{E}}, e_{ij} \notin \mathcal{E}\} \cup \{e_{ij} : e_{ij} \notin \hat{\mathcal{E}}, e_{ij} \in \mathcal{E}\}|$.

METHODOLOGY

Before introducing our new estimator GTREX, we first recall the definitions of Graphical Lasso and of neighborhood selection with Lasso. For a fixed tuning parameter $\lambda > 0$, Graphical Lasso (GLasso) estimates the precision matrix Σ^{-1} from X according to [Friedman et al., 2008]

$$\hat{\Theta}_{\text{GLASSO}}^\lambda \in \operatorname{argmin} \left\{ -\log \det(\Theta) + \operatorname{trace}(\hat{\Sigma}\Theta) + \lambda \|\Theta\|_1 \right\},$$

where the minimum is taken over all positive definite matrices $\Theta \in \mathbb{R}^{p \times p}$, $\hat{\Sigma} := X^\top X/n$ is the sample covariance matrix, and $\|\Theta\|_1 := \sum_{i,j=1}^p |\Theta_{ij}|$ is the sum of the entries of Θ . The corresponding estimator for the set of edges \mathcal{E} is then

$$\hat{\mathcal{E}}_{\text{GLASSO}}^\lambda := \{e_{ij} : |(\hat{\Theta}_{\text{GLASSO}}^\lambda)_{ij}| > 0\}. \quad (1)$$

This defines a family of graph estimators indexed by the tuning parameter λ . To assess the potential of Graphical Lasso, we define $\hat{\mathcal{E}}_{\text{GLASSO}}^* := \hat{\mathcal{E}}_{\text{GLASSO}}^{\lambda^*}$, where λ^* is the tuning parameter that minimizes the Hamming distance to the true edge set \mathcal{E} . We stress, however, that the optimal tuning parameter λ^* is *not accessible* in practice and that there are *no guarantees* that standard calibration schemes provide a tuning parameter close to λ^* . Therefore, the performance of $\hat{\mathcal{E}}_{\text{GLASSO}}^*$ is to be understood as an upper bound for the performance of Graphical Lasso.

Besides Graphical Lasso, we also consider neighborhood selection with Lasso. To this end, we define for any matrix $\bar{X} \in \mathbb{R}^{n' \times p}$, $n' \leq n$, and for any node $k \in \mathcal{V}$, the vector $\bar{X}^k \in \mathbb{R}^{n'}$ as the k th column of \bar{X} and the matrix $\bar{X}^{-k} \in \mathbb{R}^{n' \times (p-1)}$ as \bar{X} without the k th column. For a fixed tuning parameter $\lambda > 0$, the estimates of Lasso for node k are defined according to [Tibshirani, 1996]

$$\hat{\beta}_{\text{LASSO}}^\lambda(k; X) \in \operatorname{argmin}_{\substack{\beta \in \mathbb{R}^p \\ \beta_k = 0}} \left\{ \|X^k - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}. \quad (2)$$

The corresponding set of edges $\hat{\mathcal{E}}_{\text{and}}^\lambda$ (with the “and-rule”) and $\hat{\mathcal{E}}_{\text{or}}^\lambda$ (with the “or-rule”) are then defined via Algorithm 1. Similarly as above, we define the optimal representative (in terms of Hamming distance) of these families of estimators as $\hat{\mathcal{E}}_{\text{and}}^*$, called MB(and), and $\hat{\mathcal{E}}_{\text{or}}^*$, called MB(or). Again, in practice, it is not

Algorithm 1: Neighborhood selection with Lasso

Data: $X \in \mathbb{R}^{n \times p}$, $\lambda > 0$;
Result: $\hat{\mathcal{E}}_{\text{and}}^\lambda$, $\hat{\mathcal{E}}_{\text{or}}^\lambda$;
 Initialize a matrix $C := 0_{p \times p}$;
for $k = 1$ **to** p **do**
 Compute $\hat{\beta}_{\text{LASSO}}^\lambda(k; X)$ according to (2);
 Update the k th column C^k of the matrix C
 according to $C^k := \hat{\beta}_{\text{LASSO}}^\lambda(k; X)$;
end
 Set the estimated sets to
 $\hat{\mathcal{E}}_{\text{and}}^\lambda := \{e_{ij} : |C_{ij}| \vee |C_{ji}| > 0\}$ and
 $\hat{\mathcal{E}}_{\text{or}}^\lambda := \{e_{ij} : |C_{ij}| \wedge |C_{ji}| > 0\}$;

known which tuning parameters are optimal; however, MB(and) and MB(or) can highlight the potential of neighborhood selection with Lasso.

We finally introduce Graphical TREX (GTREX). To this end, we consider TREX for node k on a subsample \bar{X} according to [Lederer and Müller, 2014]

$$\hat{\beta}(k; \bar{X}) \in \underset{\substack{\beta \in \mathbb{R}^p \\ \beta_k = 0}}{\operatorname{argmin}} \left\{ \frac{\|\bar{X}^k - \bar{X}\beta\|_2^2}{\|(\bar{X}^{-k})^\top (\bar{X}^k - \bar{X}\beta)\|_\infty} + \|\beta\|_1 \right\}. \quad (3)$$

For fixed number of bootstraps $b \in \{1, 2, \dots\}$ and threshold $t \geq 0$, we then define the GTREX as the set of edges $\hat{\mathcal{E}}$ provided by Algorithm 2.

For the actual implementation, we follow Lederer and Müller [2014] and invoke that $\|a\|_\infty \approx \|a\|_q$ for q sufficiently large. We then use a projected sub-gradient methods to minimize the objective

$$\min_{\substack{\beta \in \mathbb{R}^p \\ \beta_k = 0}} \left\{ \frac{\|\bar{X}^k - \bar{X}\beta\|_2^2}{\|(\bar{X}^{-k})^\top (\bar{X}^k - \bar{X}\beta)\|_q} + \|\beta\|_1 \right\}, \quad (4)$$

which corresponds to (3).

3 EXPERIMENTAL SCENARIOS

Besides the number of parameters p , the sample size n , and the level of sparsity of the graph, the graph topology can have considerable impact on the performance of the different methods [Ravikumar et al., 2011]. For example, standard estimators require many samples for graphs with many hub nodes (nodes that are connected to many other nodes). Ravikumar et al. [2011] present a number of toy examples that confirm these theoretical predictions. The following experimental set-up is motivated by these insights. We consider six different graph topologies with varying hub structure, ranging from a Single-Hub case to Erdős-Rényi graphs:

Algorithm 2: GTREX

Data: $X \in \mathbb{R}^{n \times p}$, $b \in \{1, 2, \dots\}$, $t \in [0, 1]$;
Result: $\hat{\mathcal{E}}$, $F \in \mathbb{R}^{p \times p}$;
 Initialize all frequencies $F := 0_{p \times p}$;
for $k = 1$ **to** p **do**
 for $l = 1$ **to** b **do**
 Generate sequential bootstrap sample \bar{X} of X ;
 Compute $\hat{\beta}(k; \bar{X})$ according to (3);
 Update the frequencies for the edges adjacent to node k
 for $m = 1$ **to** p **do**
 if $m \in \operatorname{supp}(\hat{\beta}(k; \bar{X}))$ **then**
 $F_{km} := F_{km} + \frac{1}{b}$;
 end
 end
 end
 end
 Set the estimated set of edges to
 $\hat{\mathcal{E}} := \{e_{ij} : F_{ij} \vee F_{ji} > t\}$;

1. Single-Hub graph The set of edges is first set to $\mathcal{E} = \{e_{1j} : j \in \{2, \dots, p\}\}$. Until the number of edges k is exhausted, edges are then uniformly at random added to \mathcal{E} .

2. Double-Hub graph The set of edges is first set to $\mathcal{E} = \{e_{1j} : j \in \{2, \dots, p/2\}\} \cup \{e_{(p/2+1)j} : j \in \{p/2+2, \dots, p\}\}$. Until the number of edges k is exhausted, edges are then uniformly at random added to \mathcal{E} .

3. Four-Hub graph The set of edges is first set to $\mathcal{E} = \{e_{1j} : j \in \{2, \dots, p/4\}\} \cup \{e_{(p/4+1)j} : j \in \{p/4+2, \dots, p/2\}\} \cup \{e_{(p/2+1)j} : j \in \{p/2+2, \dots, 3p/4\}\} \cup \{e_{(3p/4+1)j} : j \in \{3p/4+2, \dots, p\}\}$. Until the number of edges k is exhausted, edges are then uniformly at random added to \mathcal{E} .

4. Four-niches graph Within each set of nodes $\{1, \dots, p/4\}$, $\{p/4+1, \dots, 2p/4\}$, $\{2p/4+1, \dots, 3p/4\}$, $p/4-1$ edges are uniformly selected at random and added to the set of edges. Until the number of edges k is exhausted, edges (connecting any nodes of the entire graph) are then uniformly at random added to \mathcal{E} .

5. Erdős-Rényi graph Until the number of edges k is exhausted, edges are uniformly at random added to \mathcal{E} .

6. Scale-free graph First, a set of edges is constructed with the preferential attachment algorithm [Barabási and Albert, 1999]: The set of edges is first set to $\mathcal{E} = \{e_{12}\}$. For each node $i \in \mathcal{V} \setminus \{1, 2\}$, an edge e_{ij} is then iteratively added to \mathcal{E} . The probability for selecting the edge e_{ij} is set proportional to

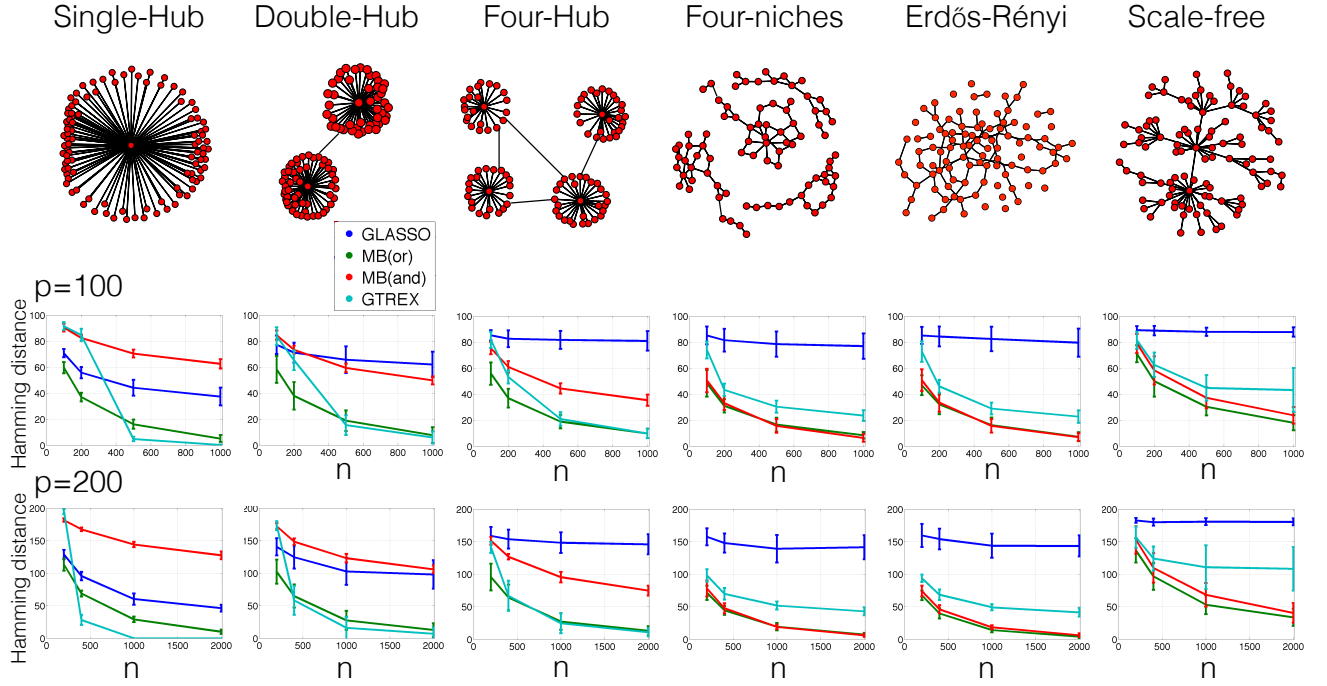


Figure 1: Hamming distances of the true graphs to GLASSO, MB(or), and MB(and) with *optimal* tuning parameter λ^* and to GTREX as a function of the sample size n . In the top row, examples of the corresponding graphs are displayed.

the degree of node $j \in \mathcal{V}$ (that is, the number of edges at node j) in the current set of edges. Until the number of edges k is exhausted, edges are then uniformly at random added to \mathcal{E} .

Given a graph \mathcal{G} that consists of a set of nodes \mathcal{V} and a set of edges \mathcal{E} as described above, a precision matrix Σ^{-1} is generated as follows: The set of edges \mathcal{E} determines which off-diagonal entries of the precision matrix Σ^{-1} are non-zero. The values of these entries are independently sampled uniformly at random in $[-a_{\max}, -a_{\min}] \cup [a_{\min}, a_{\max}]$ for some $a_{\max} > a_{\min} > 0$. The diagonal entries of Σ^{-1} are then set to a common value, which is chosen to ensure a given condition number $\text{cond} := \text{cond}(\Sigma^{-1})$ (the ratio of the largest eigenvalue to the smallest eigenvalue of Σ^{-1}).

4 NUMERICAL RESULTS

We performed all numerical computations in MATLAB 2014a on a standard MacBook Pro with 2.8GHz Dual-core Intel i7 and 16GB 1600MHz DDR3 memory. To compute the GLASSO paths, we use the C implementation of the QUIC algorithm and the corresponding MATLAB wrapper [Hsieh et al., 2011]. We set the maximum number of iterations to 200, which ensures the global convergence of the algorithm in our settings.

To compute the Lasso paths for the neighborhood selection schemes, we use the MATLAB-internal procedure `lasso.m`, which follows the popular `glmnet` R code. We implemented a neighborhood selection wrapper `mblasso.m` that returns the graph traces over the entire path for the “and-rule” and the “or-rule.” Both for GLASSO and neighborhood selection, we use a fine grid of step size 0.01 on the unit interval for the tuning parameter λ , resulting in a path over 100 values of λ . To compute TREX, we optimize the approximate TREX objective function with $q = 40$ using Schmidt’s PSS algorithm implemented in `L1General2_PSSgb.m`. We use the PSS algorithm with the standard parameter settings and set the initial solution to the parsimonious all-zeros vector $\beta_{\text{init}} = (0, \dots, 0)^T \in \mathbb{R}^p$. We use the following PSS stopping criteria: minimum relative progress tolerance `optTol`=1e-7, minimum gradient tolerance `progTol`=1e-9, and maximum number of iterations `maxIter` = 0.2p. We implemented a wrapper `gtrex.m` that integrates the node-wise TREX solutions and returns the frequency table for each edge and the resulting graph estimate. We use $b = 31$ bootstrap samples in B-TREX; increasing the number of bootstraps did not result in significant changes of the GTREX solutions. The generation of the graphs and precision matrices is implemented in our new MATLAB toolbox GMG (Graphical Model Genera-

Table 1: **Precision** and **Recall** for Single-Hub graph with $a_{\min} = 0.2$, $a_{\max} = 1$, $k = p - 1$, and $\text{cond} = 100$.

$n = 100, p = 100$			$n = 500, p = 100$		
Method	P	R	Method	P	R
GLASSO	0.99	0.35	GLASSO	0.99	0.59
MB(or)	0.99	0.48	MB(or)	0.99	0.87
MB(and)	0.99	0.49	MB(and)	0.99	0.92
GTREX	0.99	0.13	GTREX	1.00	0.99

$n = 200, p = 200$			$n = 1000, p = 200$		
Method	P	R	Method	P	R
GLASSO	1.00	0.25	GLASSO	1.00	0.44
MB(or)	1.00	0.30	MB(or)	1.00	0.58
MB(and)	1.00	0.29	MB(and)	1.00	0.59
GTREX	0.99	0.05	GTREX	1.00	0.60

 Table 2: **Precision** and **Recall** for Four-Hub graph with $a_{\min} = 0.2$, $a_{\max} = 1$, $k = p - 1$, and $\text{cond} = 100$.

$n = 100, p = 100$			$n = 500, p = 100$		
Method	P	R	Method	P	R
GLASSO	1.00	0.17	GLASSO	0.99	0.19
MB(or)	1.00	0.55	MB(or)	1.00	0.86
MB(and)	0.99	0.59	MB(and)	1.00	0.93
GTREX	0.99	0.26	GTREX	1.00	0.80

$n = 200, p = 200$			$n = 1000, p = 200$		
Method	P	R	Method	P	R
GLASSO	1.00	0.15	GLASSO	0.99	0.17
MB(or)	1.00	0.36	MB(or)	1.00	0.54
MB(and)	1.00	0.40	MB(and)	1.00	0.57
GTREX	1.00	0.20	GTREX	1.00	0.53

tor), which will be made available on the authors' web-sites.

We generate the graphical models as outlined in Section 3. We set the number of nodes to $p \in \{100, 200\}$, the number of edges to $k = p - 1$, the bounds for the absolute values of the off-diagonal entries of the precision matrix to $a_{\min} = 0.2$ and $a_{\max} = 1$, and the condition number to $\text{cond} = 100$. We then draw $n \in \{p, 2p, 4p, 10p\}$ samples from the resulting normal distribution and normalize each sample to have Euclidean norm equal to \sqrt{n} . We measure the performance of the estimators in terms of Hamming distance to the true graph and in terms of Precision/Recall. We stress that for GLASSO, MB(or), and MB(and), we select the (in practice unknown) tuning parameter λ that minimizes the Hamming distance to the true graph. For GTREX, we set the frequency threshold to $t = 0.75$; however, it turns out that GTREX is robust with respect to the choice of the threshold. For each graph, we report the averaged results over 20 repetitions.

 Table 3: **Precision** and **Recall** for Erdős-Rényi graph with $a_{\min} = 0.2$, $a_{\max} = 1$, $k = p - 1$, and $\text{cond} = 100$.

$n = 100, p = 100$			$n = 500, p = 100$		
Method	P	R	Method	P	R
GLASSO	1.00	0.31	GLASSO	0.99	0.42
MB(or)	1.00	0.61	MB(or)	1.00	0.89
MB(and)	0.99	0.69	MB(and)	1.00	0.94
GTREX	0.99	0.36	GTREX	1.00	0.71

$n = 200, p = 200$			$n = 1000, p = 200$		
Method	P	R	Method	P	R
GLASSO	1.00	0.30	GLASSO	0.99	0.42
MB(or)	1.00	0.43	MB(or)	1.00	0.57
MB(and)	1.00	0.46	MB(and)	1.00	0.58
GTREX	1.00	0.34	GTREX	1.00	0.45

The results are summarized in Figure 1 and in Tables 1, 2, and 3. The results with respect to Hamming distance in Figure 1 provide three interesting insights: First, GLASSO performs poorly in Hamming distance for all considered scenarios. We suspect that this is connected with the chosen value for the condition of the precision matrix. Second, we observe marked differences between MB(and) and MB(or). In particular, the two methods have a similar performance in the scenarios with the Four-niches and the Erdős-Rényi graphs but a completely different performance in the scenarios with the hub graphs. Third, GTREX performs excellently for the hub graphs if the sample size is sufficiently large ($n > 2p$) and reasonably in all other scenarios. The results with respect to Precision/Recall in Tables 1, 2, 3 show that all methods have an excellent Precision (all values are close to 1), but differ in Recall. MB(and) provides the best overall performance, while GTREX is especially competitive in the scenarios with larger sample size n .

TREX does not contain a tuning parameter, but one can argue that the frequency threshold t could be adapted to the model or the data and, therefore, plays the role of a tuning parameter in GTREX. However, the above results demonstrate that the universal value $t = 0.75$ works for a large variety of scenarios. Moreover, GTREX is robust with respect to the choice of t . This is illustrated in Figure 2, where for two scenarios, we report the Hamming distances of GTREX to the true graphs as a function of t . We observe that the Hamming distances are similar over wide ranges of t . In the same figure, we also report the Hamming distances of the standard methods to the true graphs as a function of the tuning parameter λ . We see that these paths have narrow peaks, which suggests that the tuning parameters of GLASSO and of neighborhood selection with Lasso need to be carefully calibrated.

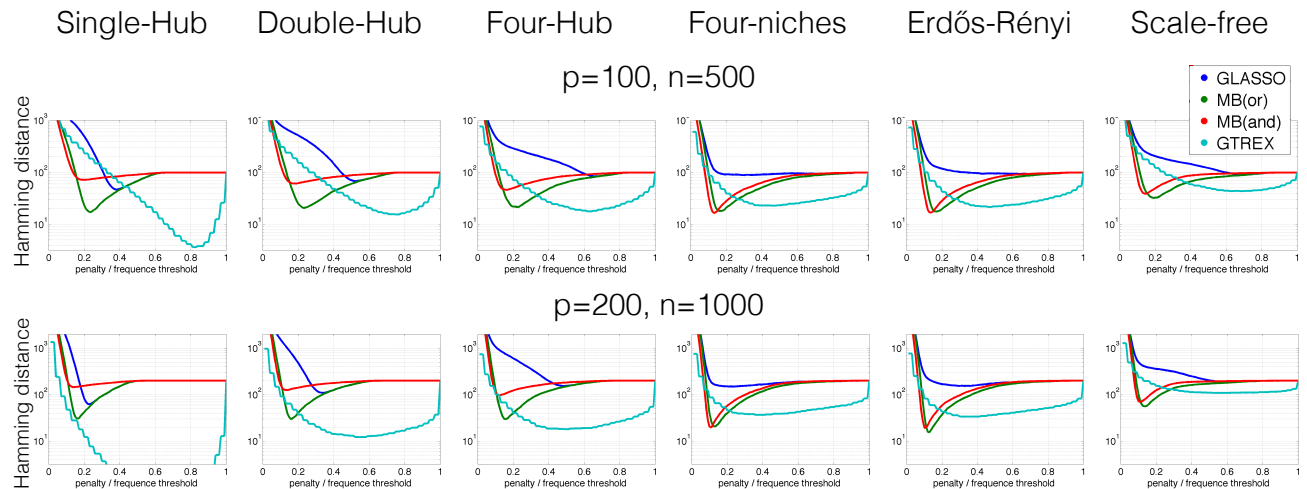


Figure 2: Paths over the tuning parameter λ in GLASSO, MB(or), and MB(and) and paths over the threshold t in GTREX.

5 Conclusions

Our simulations reveal that the potential of Graphical Lasso and the potential of neighborhood selection depend differently on the graph topology. On the other hand, the simulations also indicate that neighborhood selection with the “or-rule” and optimal tuning parameter provides accurate graph estimation across a large variety of scenarios.

In practice, it is unknown which tuning parameter is optimal, and the calibration of tuning parameters can be challenging. We have, therefore, introduced Graphical TREX (GTREX), which makes parameter tuning obsolete. Our simulations demonstrate that, in many scenarios, GTREX rivals MB(or) in terms of Hamming distance to the true graph and Precision/Recall. This suggests that GTREX can provide accurate graph estimation for a variety of graph types without requiring the calibration of a tuning parameter.

A constant threshold parameter of size 0.75 works well for GTREX in all considered scenarios. In addition, the paths over the threshold parameter are typically flat, which demonstrates that GTREX is robust with respect to the choice of this parameter. In future work, a refined choice (incorporating, for example, the sample size n and the number of parameters p) could be explored.

A next step is to compare GTREX with combinations of standard methods and schemes for the calibration their tuning parameter, see, for example, [Liu et al., 2010, Liu and Wang, 2012].

References

- Onureena Banerjee, Laurent El Ghaoui, and Alexandre D’Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516, 2008.
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- Nir Friedman. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science*, 303(5659):799–805, 2004.
- Cho-Jui Hsieh, Matyas Sustik, Inderjit Dhillon, and Pradeep Ravkumar. Sparse inverse covariance matrix estimation using quadratic approximation. *NIPS*, pages 1–18, 2011.
- David Jones, Daniel Buchan, Domenico Cozzetto, and Massimiliano Pontil. PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics*, 28:184–190, 2012.
- Zachary Kurtz, Christian Müller, Emily Miraldi, Dan Littman, Martin Blaser, and Richard Bonneau. Sparse and compositionally robust inference of microbial ecological networks. *arXiv preprints*, 2014.
- Clifford Lam and Jianqing Fan. Sparsistency and Rates of Convergence in Large Covariance Matrix Estimation. *Annals of Statistics*, 37(6B):4254–4278, 2009.

- Steffen Lauritzen. *Graphical models*. Oxford University Press, 1996.
- Johannes Lederer and Christian Müller. Don't fall for tuning parameters: Tuning-free variable selection in high dimensions with the TREX. *arXiv preprint*, 2014.
- Han Liu and Lie Wang. Tiger: A tuning-insensitive approach for optimally estimating gaussian graphical models. *arXiv preprint*, 2012.
- Han Liu, John Lafferty, and Larry Wasserman. The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. *J. Mach. Learn. Res.*, 10:2295–2328, 2009.
- Han Liu, Kathryn Roeder, and Larry Wasserman. Stability approach to regularization selection (stars) for high dimensional graphical models. *NIPS*, 2010.
- Qiang Liu and Alexander Ihler. Learning scale free networks by reweighted l1 regularization. *International Conference on Artificial Intelligence and Statistics*, 2011.
- Nicolai Meinshausen and Peter Bühlmann. High Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.
- Pradeep Ravikumar, Martin Wainwright, John Lafferty, and Others. High-dimensional Ising model selection using L1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.
- Pradeep Ravikumar, Martin Wainwright, Garvesh Raskutti, and Bin Yu. High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.
- Anja Wille, Philip Zimmermann, Eva Vranová, Andreas Fürholz, Oliver Laule, Stefan Bleuler, Lars Hennig, and Others. Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol*, 5(11):R92, 2004.
- Ming Yuan and Yi Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1): 19–35, 2007.